EG&G ROCKY FLATS

INTEROFFICE CORRESPONDENCE

DATE: September 10, 1993

TO: D. M. Smith, Environmental Remediation Management

FROM: D. R. Weier, Ph.D./A. D. Palachek, Ph.D., Stat Apps, Bldg. 850, X4194/7973

SUBJECT: COMMENTS ON HELSEL PAPER AND STATISTICS METHODOLOGY RECOMMENDATIONS DRW-021-93

This memo is provided per your request at the meeting this morning between you, Denny Weier, Mary Siders, and Randy Boan. It is in regard to the Helsel paper which appears in Environmental Science and Technology (1990), Vol. 24, No. 12. In this paper it is suggested that the nonparametric scores tests are inappropriate for unequal sample sizes.

Helsel refers to the Latta paper which appears in the Journal of the American Statistical Association (1981), Vol. 76, No. 375, and Helsel states that in this reference "most of the score tests were found inappropriate for the case of unequal sample sizes." Such a statement cannot be found anywhere in the Latta paper, and Helsel gives no further explanation for what is apparently his own conclusion.

We disagree with Helsel's statement. Latta shows, through simulation studies, that differing censoring mechanisms between the two samples tend to result in a small increase to the Type I error rate. The amount of increase is difficult to ascertain, as Latta only presented results for two cases: (i) same censoring mechanism, and (ii) one sample censored and no censoring in the other sample. Latta's presented results show no effect of unequal sample size on case (i), and a slight Type I error increase for case (ii). An increase in Type I error will result in greater power for the test, that is, the increased likelihood of identifying potential contaminants of concern.

Latta only examined scores tests and did not compare them to any other testing approach such as t-tests after replacement of censored data. We feel that a simulation that would include alternative test approaches would indicate that these alternatives are more dramatically affected by different censoring mechanisms and unequal sample sizes than the nonparametric scores tests.

Latta does state that "unequal sample sizes rule out the use of the conditional permutation variance." However, this is not the variance estimator used in the Statistical Applications recommended approach. Latta's final recommendation is that "with heavy censoring and sample sizes that are far apart or censoring mechanisms that differ greatly, the Peto-Prentice statistic with the asymptotic variance should be used." This Peto-Prentice statistic is one that is very similar to the Gehan statistic. It differs slightly in the way in which the ranks are assigned to data.

Statistical Applications recommended the use of the Gehan test over the Peto-Prentice primarily due to the more intuitive appeal of the Gehan ranks and thus the easier explanation to non-statistical personnel. The difference in performance of the two tests is not at all dramatic. The Peto-Prentice can be used instead of the Gehan test if desired. Both Dr.'s Gilbert and Crump would likely agree with the continued use of the Gehan test or the switch to the Peto-Prentice test.

In summary, the simulation results presented in the literature are fairly scant and show only minor differences between various scores tests. While statisticians often recommend tests based on such minor differences, these results give no reason to make claims that any method is inappropriate. Much more detailed studies would need to be performed to support a statement regarding the appropriateness of a particular test as in Helsel's statement.

In light of this discussion, Gilbert's recommendations, and Crump's review comments on Gilbert's recommendations, Statistical Applications personnel recommend the following analysis steps:

1. Graphical summaries of data, specifically boxplots for radionuclides and "circle/plus" plots for analytes containing nondetects. Examples of these have been previously provided.

2. "Hot Measurement" tests preferably based on non-statistically derived standards. If this is impractical, 99/99 UTL's should be used rather than the 95/95 UTL's.

3. The application of a nonparametric scores test. The Gehan or Peto-Prentice are appropriate.

Contact us with questions or for additional information.